MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **82** (2019) 205-218

ISSN 0340 - 6253

Dynamic Representations of Biological Sequences

Dorota Bielińska–Waż[a], Damian Panas[a], Piotr Waż[b]

[a]*Department of Radiological Informatics and Statistics, Medical University of Gdańsk, Gdańsk, Poland*
djwaz@gumed.edu.pl, dpanas@gumed.edu.pl

[b]*Department of Nuclear Medicine, Medical University of Gdańsk, Gdańsk, Poland*
phwaz@gumed.edu.pl

(Received January 6, 2019)

Abstract

Methods of bioinformatics in which the biological sequences (DNA, RNA, protein) are represented by sets of material points in 2D, 3D, or 20D space, and described by values analogous to the ones used in the dynamics, as e.g. moments of inertia, are reviewed. A new application of the 3D method, called by us *3D-Dynamic Representation of DNA/RNA Sequences* is proposed. It is shown that the method is useful for a description of complete genome sequences of dengue virus.

1 Introduction

In 2007, we developed a new bioinformatics method and invented its descriptive name *2D-Dynamic Representation of DNA/RNA Sequences*. At present the name refers to a class of bioinformatics methods, but originally it has been assigned to the two-dimensional version only [1–3]. The method has been extended to three-dimensions (*3D-Dynamic Representation of DNA/RNA Sequences*) [4] and a twenty-dimensional version of the method (*20D-Dynamic Representation of Protein Sequences*) has also been designed [5]. In the present work, we demonstrate the practical power and usefulness of this approach by its application to a description of the complete genome sequences of dengue virus.

The main idea of the *Dynamic Representations of Biological Sequences* (DRBS) is transforming the sequences to sets of material points in 2D, 3D, or 20D space. We call these methods *dynamic*, because we treat the biochemical objects as rigid bodies and characterize them by numerical characteristics (*descriptors*), analogous to the ones used in the classical dynamics as, for example, moments of inertia.

In this paper, after reviewing in Section 2 the present status of the DRBS methods, we discuss, in Section 3, an application of the 3D version of the method to the description of the dengue virus genome.

2 DRBS Methods

In the DRBS methods the biological sequences (DNA, RNA, protein) are represented by *dynamic graphs*. The graphs are composed of point-masses in a properly defined Euclidean space. The shape and the location of a dynamic graph in the space is characteristic for the sequence described by this graph. So far, DRBS methods utilizing dynamic graphs defined in 2D and 3D spaces (applied to DNA/RNA sequences) [1, 4] and in 20D space (applied to protein sequences) [5] have been designed.

The DNA/RNA sequence is a sequence composed of four letters A, C, G and T/U corresponding to four bases – Adenine, Cytosine, Guanine, and Thymine/Uracil, respectively. In DBRS methods each base is represented by a unit basis vector.

A dynamic graph is defined by the method of its construction. The graph is developed recursively, performing a series of shifts (walks) along the basis vectors representing the bases, starting from the origin of the Cartesian coordinate system in the pertinent Euclidean space. Techniques using the idea of shifts (or walks) have been introduced as the first graphical representations of the DNA sequences – in three dimensions by Hamori and Ruskin [6] and in two dimensions by Gates [7], Nandy [8], Leong and Morgenthaler [9]. In two dimensions, the three mentioned above methods use four orthogonal directions for the representations of the four bases. They differ by the assignment of the basis vectors to particular bases. The most popular diagrams are the Nandy plots [8]. The two-dimensional methods are the most convenient techniques for the visualization. But, in their original form, the methods are not free from non-uniqueness, referred to as *degeneracy* (in some cases different sequences are represented by the same diagrams). In order to reduce or to remove the degeneracy, many new Graphical Representation (GR) methods have been designed, including the DRBS ones. One of the most important features of the graphical representations of DNA/RNA sequences is their usefulness in both graphical and numerical revealing different aspects of similarity of the objects they describe [10–26] (for reviews see [27, 28]).

The range of applicability of the GR methods have been also extended to a description of protein sequences, e.g. [29, 30]. A numerical representation of proteins as walks in a 20D space, introduced by Novič and Randić [31], has been applied in 20D DRBS [5].

The DRBS graph is composed of a set mass-one point masses. The biological sequence is mapped onto the distribution of the point masses forming the dynamic graph. The graph is constructed by properly locating a point-mass at the end of each vector representing a base. In particular, in the 2D-Dynamic Representation of DNA/RNA Sequences the bases are represented by the basis vectors, the same as in the Nandy plots: A = $(-1, 0)$, G = $(1, 0)$, C = $(0, 1)$ and T/U = $(0, -1)$. If, for example, the first base is G, then the vector representing the first shift is $(1, 0)$ and the coordinates of the first point-mass are $(1, 0)$. This point-mass is a starting point for the second shift representing the second base in the sequence. At the end of the second vector the second point-mass is located. The procedure is repeated until the end of the sequence. If the ends of some vectors meet several times at the same point, then the mass of this point increases: it is the sum of masses corresponding to all vectors ending at this point. In the Nandy plots, if the shifts are performed back and forth along the same trace then some parts of the sequence are hidden. As a consequence, different sequences may be represented by the same plot. In the 2D-Dynamic Representation of DNA/RNA Sequences, by summing up masses at the points visited by several vectors, this kind of degeneracy is removed. A drawback of this approach is that the history of emergence of a graph is lost – the graphs self-overlap. This feature has been corrected in the 3D-Dynamic Representation of DNA/RNA Sequences. In the 3D method we represent the bases as follows: A = $(-1, 0, 1)$, G = $(1, 0, 1)$, C = $(0, 1, 1)$, and T/U = $(0, -1, 1)$. The construction of the dynamic graph is here analogous to the one in the 2D method. The shifts start from the origin of the 3D coordinate system and continue using subsequent bases in the sequence. Similarly as in the 2D case, to obtain the dynamic graph, at the end of each vector we locate a point-mass with mass equal to 1.

For the representation of protein sequences we choose a 20D space. The assignment of the axes of the 20-dimensional coordinate system to the amino acids has been described in [5]. The construction of a 20D-dynamic graph is the same as in the 2D and the 3D methods, i.e. we start shifts at the origin of the 20D coordinate system, and at the end of the vectors we locate the point-mass with mass equal to 1. As a consequence of the choice of the basis vectors, the masses of all the points in the 3D and in the 20D method are equal to 1. The 20D method for protein sequences can easily be extended to more dimensions, if more than 20 amino acids were present in the sequence. New basis vectors can be introduced in the same way as it has been done for 20 dimensions, and the upper limit in all the summations present in the method can be modified accordingly.

For a numerical characterization of the 2D-dynamic graphs we have selected the coordinates of the center of mass and the principal moments of inertia of the graph in the 2D space [1]. Using the 2D-dynamic graphs we have created the corresponding mass-density distributions and took moments of these distributions as descriptors [2]. We also have considered the angles between the x axis and the principal axis of the 2D-dynamic graph as descriptors. Good descriptors are also coordinates of the center of mass divided by the principal moments of inertia [32]. We have also described the 2D-dynamic graphs using matrix elements of the tensor of inertia and the graph radius, i.e. the length of the position vector of the mass center [33].

In the case of the 3D-dynamic graphs, analogously as in the 2D method, for the characterization of the graphs we have used coordinates of the center of mass, and the principal moments of inertia in the 3D space [4]. We have also used as descriptors the coordinates of the center of mass divided by the principal moments of inertia or divided by the normalized principal moments of inertia, and the cosines of the angles between different planes [4].

For the characterization of the 20D-dynamic graphs representing the protein sequences, we have chosen the 20D normalized principal moments of inertia, and the sum of all 20 normalized principal moments of inertia [5].

The 2D and the 3D methods allow for a direct visualization of the objects. In order to visualize the 20D graphs we have projected them onto 2D or 3D spaces.

In order to study the similarity/dissimilarity between sequences one has to define a similarity measures. In particular, for comparing sequences we have introduced a new normalized similarity measure [34]. We have also introduced mass overlaps as another similarity measure [3]. In similarity studies, the 2D-dynamic graphs have been treated as rigid bodies. In order to find the maximum of the overlap of masses of a pair of graphs, the graphs are shifted and rotated. The numerical procedures are based on the genetic algorithms [3].

It is worth to notice, that the idea of characterizing biological sequences by the moments of inertia, introduced by us [1], has already been adopted by other authors. Yao *et al.* applied this idea for the representation of protein sequences by 2D moments of inertia [35] and by 3D moments of inertia [36]. Hou et al. applied 3D moments of inertia to characterize graphs representing protein sequences [37].

The method introduced by us, 2D-Dynamic Representation of DNA Sequences, has also been generalized to three dimensions by Aram and Iranmanesh [38]. As a consequence, two different 3D methods derived from the same method and having the same name are present in the literature [4, 38].

In the present work, for the characterization of the sequences we apply the 3D-Dynamic Representation of DNA/RNA Sequences in the sense introduced earlier by us: the sequences are represented by the material points and described by values used in classical dynamics [4]. The following descriptors are used in this work:

- Coordinates of the centers of mass of the 3D-dynamic graphs μ_a, where $a = x, y, z$;
- Matrix elements of the tensor of the moment of inertia of the 3D-dynamic graphs I_{ab}, where $b = x, y, z$;
- Values $D_k^a = \mu_a / I_k$, where I_k are the principal moments of inertia of the 3D-dynamic graphs and $k = 1, 2, 3$.

The coordinates of the center of mass of the 3D-dynamic graph, in the XYZ coordinate system are defined as

$$\mu_a = \frac{\sum_{i=1}^{N} m_i a_i}{\sum_{i=1}^{N} m_i}, \tag{1}$$

where $(x_i,\, y_i,\, z_i)$ are the coordinates of the mass m_i. We assume $m_i = 1$ for all the points. Then, the length of the sequence N is equal to the total mass of the 3D-dynamic graph

$$N = \sum_{i=1}^{N} m_i \tag{2}$$

and the coordinates of the center of mass of the 3D-dynamic graph are

$$\mu_a = \frac{1}{N} \sum_{i=1}^{N} a_i. \tag{3}$$

The tensor of the moment of inertia is defined by the matrix

$$\hat{I} = \begin{pmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{pmatrix} \tag{4}$$

with the matrix elements

$$I_{aa} = \sum_{i=1}^{N} m_i \left[(b_i')^2 + (c_i')^2 \right], \quad I_{ab} = I_{ba} = -\sum_{i=1}^{N} m_i a_i' b_i', \tag{5}$$

where $\{a, b, c\} = \{x, y, z\}$, $a \neq b \neq c$ and $(x_i',\ y_i',\ z_i')$ are the coordinates of m_i in the Cartesian coordinate system with the origin selected at the center of mass. The *principal moments of inertia* I_k are defined as the eigenvalues of $\hat{I}$

$$\hat{I}\omega_k = I_k \omega_k, \tag{6}$$

where ω_k are the eigenvectors of the problem.

The eigenvalues are obtained by solving the following set of equations:

$$\begin{vmatrix} I_{xx} - I & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} - I & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} - I \end{vmatrix} = 0. \tag{7}$$

The eigenvectors $\omega_1, \omega_2, \omega_3$ form the basis for a new coordinate system. The corresponding axes of this new system are denoted Ω_1, Ω_2, Ω_3 and referred to as *the principal axes*. The eigenvalues I_1, I_2, I_3, are called the *principal moments of inertia* and are equal to the moments of inertia associated with the rotations around the principal axes.

The moment of inertia of an object about a rotational axis describes how difficult is to induce the rotation of the object around this axis. If the mass is concentrated close to the axis, it is easier to accelerate into spinning fast and the moment of inertia is smaller. As a consequence, the values of these descriptors reflect the concentrations of masses around the axes.

In our previous work we also used the square roots of the normalized principal moments of inertia [4]:

$$r_1 = \sqrt{\frac{I_1}{N}}, \quad r_2 = \sqrt{\frac{I_2}{N}}, \quad r_3 = \sqrt{\frac{I_3}{N}}. \tag{8}$$

The relative orientation of the new and old coordinate system may be described by the cosines of properly defined angles. Let M_1, M_2 and M_3 denote, respectively, the planes (X, Y), (X, Z) and (Y, Z). Similarly, Q_1, Q_2, Q_3 stand for the planes (Ω_1, Ω_2), (Ω_1, Ω_3),

(Ω_2, Ω_3), respectively. For the characterization of the 3D-dynamic graphs we used the cosines of the angles between the planes of the two systems of coordinates (Fig. 1) [4]:

$$C_{ij} \equiv \cos\left(M_i, Q_j\right), \quad i, j = 1, 2, 3. \tag{9}$$

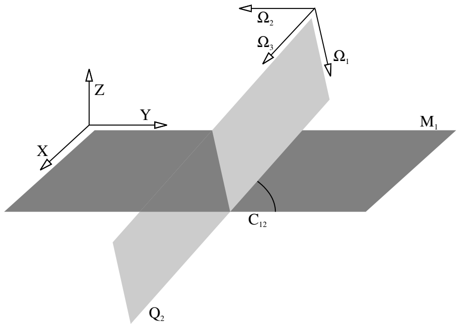

Figure 1. Planes M_1:(X,Y), M_2:(X,Z), M_3:(Y,Z), Q_1:(Ω_1,Ω_2), Q_2:(Ω_1,Ω_3), Q_3:(Ω_2,Ω_3).

3 Results and Discussion

Recently we applied the 2D version of the DRBS approach to the description of Zika [33] and influenza [39] virus genomes. In the present work, complete strains of the dengue virus genome are studied using the 3D method. Each year dengue virus (DENV) causes dengue fever in about 390 million individuals in over 100 countries [40–42]. DENV exists in four serotypes (DENV-1 – DENV-4).[1]

In the present calculations, 2712 sequences have been used: 822 sequences of serotype 1 (DENV-1), 827 of serotype 2 (DENV-2), 829 of serotype 3 (DENV-3), and 234 of serotype 4 (DENV-4). The representative examples of the results (the principal moments of inertia) for 25 sequences for each serotype are listed in Table 1. The complete set of results is available from the authors upon request.

The distribution of the x coordinate of the center of mass of the 3D-dynamic graphs (μ_x) representing all 2712 sequences is visualized in Fig. 2. As one can see, the distribution is composed of 3 separate sub-distributions. The values of μ_x cluster around 3 values.

[1] In 2013 the discovery of the fifth serotype was reported: Normile, D. (2013) Tropical medicine. Surprising new dengue virus throws a spanner in disease control efforts, Science. 342: 415. DOI: 10.1126/science.342.6157.415.

Table 1. Selected sequence data and the principal moments of inertia.

Serotype 1						Serotype 2					
No.	Accession	N	$I_1/10^{10}$	$I_2/10^{10}$	$I_3/10^6$	No.	Accession	N	$I_1/10^{10}$	$I_2/10^{10}$	$I_3/10^6$
1	LC335877	10179	8.823	8.822	10.32	26	AF489932	10722	10.34	10.34	5.255
2	LC335871	10179	8.823	8.822	9.932	27	AJ487271	10597	9.979	9.979	4.361
3	LC335872	10179	8.823	8.822	10.16	28	LC367234	10723	10.33	10.33	5.096
4	LC335873	10179	8.823	8.822	10.02	29	FJ390389	10632	10.07	10.07	6.343
5	LC335879	10179	8.823	8.822	10.41	30	AJ968413	10723	10.34	10.34	4.236
6	AB608789	10677	10.18	10.18	9.393	31	AB543624	10731	10.37	10.37	2.385
7	AB608788	10693	10.23	10.23	10.64	32	AB189122	10723	10.33	10.33	4.123
8	AB608786	10749	10.39	10.39	9.838	33	AB189123	10723	10.33	10.33	4.252
9	AB608787	10749	10.39	10.39	11.13	34	AB189124	10723	10.33	10.33	4.311
10	HE795086	10179	8.826	8.825	9.930	35	AB122020	10723	10.34	10.34	6.044
11	AB519681	10735	10.35	10.35	5.922	36	AB122022	10723	10.34	10.34	5.942
12	AB189120	10735	10.35	10.35	5.904	37	LC129169	10723	10.34	10.34	3.162
13	AB189121	10735	10.35	10.35	7.190	38	LC121816	10685	10.22	10.22	4.539
14	AB195673	10718	10.30	10.30	6.101	39	AB479041	10647	10.11	10.11	6.136
15	AB204803	10706	10.27	10.27	6.357	40	LC111438	10678	10.21	10.21	3.842
16	LC011948	10693	10.23	10.23	11.30	41	KF744400	10176	8.832	8.832	6.189
17	LC016760	10693	10.23	10.23	11.10	42	KF744401	10176	8.832	8.832	6.138
18	LC128301	10693	10.23	10.23	9.953	43	KF744402	10176	8.834	8.834	5.044
19	LC011945	10693	10.23	10.23	11.12	44	KY474330	10648	10.13	10.13	5.436
20	LC011949	10693	10.23	10.23	11.54	45	KY474334	10460	9.602	9.601	5.137
21	HM469966	10735	10.35	10.35	12.47	46	KY474325	10648	10.13	10.13	5.601
22	JX669462	10737	10.35	10.35	6.281	47	KY474324	10648	10.13	10.13	5.535
23	JX669463	10735	10.35	10.35	5.948	48	KY474318	10648	10.13	10.13	5.527
24	JX669464	10736	10.35	10.35	6.809	49	KY474315	10648	10.13	10.13	5.496
25	JX669465	10736	10.36	10.36	7.453	50	KY474316	10647	10.13	10.13	6.654
Serotype 3						Serotype 4					
No.	Accession	N	$I_1/10^{10}$	$I_2/10^{10}$	$I_3/10^6$	No.	Accession	N	$I_1/10^{10}$	$I_2/10^{10}$	$I_3/10^6$
51	LT898451	10707	10.28	10.28	7.102	76	KY474335	10642	10.07	10.07	12.55
52	LT898452	10707	10.28	10.28	6.848	77	KR922405	10164	8.769	8.768	10.13
53	AB189125	10707	10.28	10.28	7.118	78	KP188557	10425	9.467	9.466	13.51
54	AB189126	10707	10.28	10.28	7.210	79	KP188558	10572	9.873	9.872	12.08
55	AB189127	10707	10.28	10.28	7.321	80	KP188560	10576	9.884	9.883	12.11
56	AB189128	10707	10.27	10.27	5.750	81	KP188562	10649	10.09	10.09	12.11
57	KY794787	10617	10.02	10.02	7.380	82	KP188563	10654	10.11	10.10	12.12
58	KY794788	10590	9.945	9.945	7.588	83	KP188564	10618	10.00	10.00	13.80
59	KY794789	10634	10.07	10.07	7.540	84	KP188566	10426	9.471	9.470	11.56
60	KY794790	10634	10.07	10.07	7.666	85	JN983813	10649	10.09	10.09	12.73
61	KY794786	10643	10.10	10.10	8.330	86	MG272272	10652	10.10	10.10	12.83
62	JX669489	10709	10.28	10.28	8.546	87	MG272274	10652	10.10	10.10	11.96
63	JX669498	10707	10.28	10.28	8.109	88	KF041260	10652	10.10	10.09	15.06
64	KF954946	10686	10.21	10.21	6.968	89	FJ024476	10606	9.970	9.968	14.63
65	KF954947	10675	10.18	10.18	7.054	90	GQ868579	10593	9.932	9.931	14.05
66	KF954948	10678	10.19	10.19	7.187	91	GQ868580	10598	9.946	9.945	11.41
67	KF954949	10677	10.19	10.19	7.076	92	GQ868581	10592	9.929	9.929	11.41
68	KF954945	10686	10.21	10.21	7.250	93	GQ868582	10606	9.969	9.968	11.43
69	JX669495	10707	10.28	10.27	7.997	94	GQ868583	10551	9.814	9.813	11.03
70	JX669499	10707	10.28	10.28	7.715	95	GQ868584	10552	9.817	9.816	11.19
71	KY863456	10707	10.28	10.28	7.414	96	GQ868585	10606	9.970	9.969	14.51
72	JX669492	10709	10.28	10.28	8.423	97	JQ922560	10739	10.34	10.34	13.04
73	JX669500	10707	10.28	10.28	8.180	98	JQ922558	10626	10.03	10.02	12.56
74	JX669497	10709	10.28	10.28	7.798	99	JQ922559	10616	9.992	9.991	10.21
75	JX669501	10707	10.28	10.28	8.953	100	KP406806	10664	10.13	10.13	13.48

The left sub-distribution, with the smallest values of μ_x, is formed by serotype 2, the right sub-distribution is formed by serotype 4, and the middle one by both 1 and 3 serotypes.

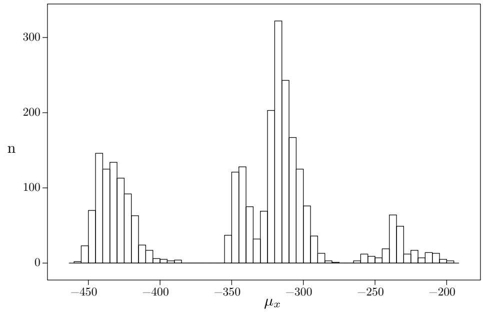

Figure 2. Distribution of μ_x for 2712 3D-dynamic graphs representing complete genome sequences of dengue virus.

Examples of 3D-dynamic graphs are shown in Fig. 3. The sequence of serotype 1 is there compared with other serotypes. Two sequences of the same serotype (top, left panel) are represented by similar graphs. The differences between sequences are graphically clearly visible: serotypes 1 and 2 (top, right panel) and serotypes 1 and 4 (bottom, right panel). The differences between the graphs representing serotypes 1 and 3 (bottom, left panel) are smaller.

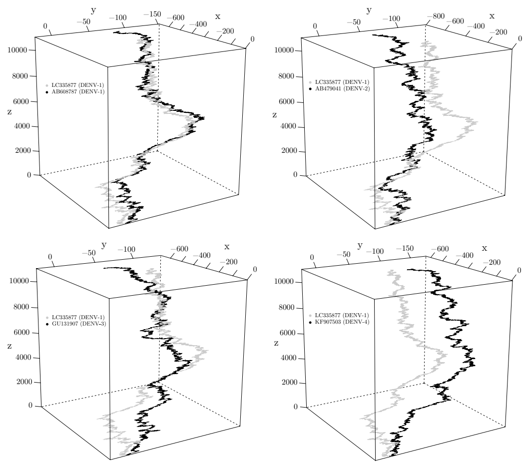

Figure 3. 3D-dynamic graphs representing the complete genome sequences of dengue virus.

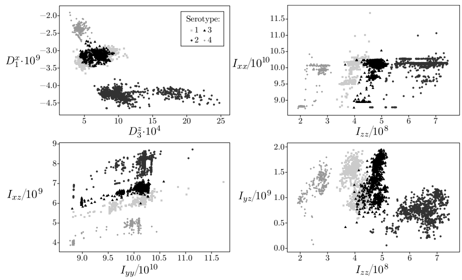

Figure 4. Two-dimensional classification diagrams.

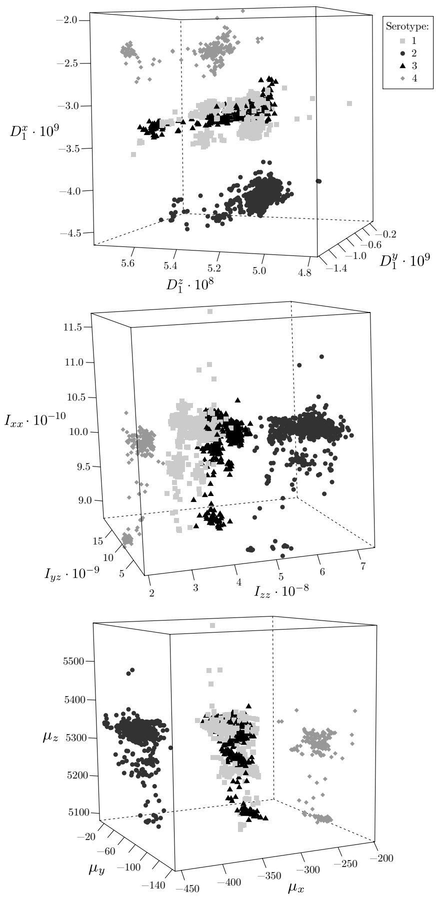

Figure 5. Three-dimensional classification diagrams.

The classification of all 2712 sequences is clearly seen in two- and three-dimensional classification diagrams (Figs. 4 and 5, respectively). In the two-dimensional diagrams (Fig. 4) relations between pairs D_1^x and D_3^z (top, left panel), I_{xx} and I_{zz} (top, right panel), I_{xz} and I_{yy} (bottom, left panel), I_{yz} and I_{zz} (bottom, right panel) are displayed. Similarly, in the three-dimensional diagrams (Fig. 5) relations between triples of quantities $D_1^x - D_1^z - D_1^y$ (top panel), $I_{xx} - I_{yz} - I_{zz}$ (middle panel), and $\mu_z - \mu_y - \mu_x$ (bottom panel) are displayed. Different serotypes are represented by different symbols in the plots. The symbols are the same in all diagrams. Similarly as in the images presented in Figures 2 and 3, also here, serotypes 1 and 3 cluster, and as a consequence we observe three separate groups: serotypes 1 and 3 (group 1), serotype 2 (group 2), and serotype 4 (group 3).

Our results (clustering to 3 groups) are consistent with the known antigenic relationships between DENV serotypes: DENV-1 and DENV-3 share common epitopes that are not present in DENV-2 or DENV-4 [43].

In the present work, we used in the calculations a large number of sequences (several thousands). As the result, the sequences have been correctly classified. As one can see, the descriptors characterizing the sequences are located in different parts of the classification diagrams (Figs. 4 and 5). Additionally, the 3D-dynamic graphs proved to be a simple intuitive tool which may easily be applied to a graphical comparison of the sequences (Fig. 3). Consequently, this approach may be used in both numerical and graphical studies of biomedical sciences. Therefore, the creation of a public server available online for the analysis of the sequences based on the DRBS-based approaches is an important aim of our future work.

Summarizing, the present results supply another confirmation of the usefulness of DRBS-based approaches to the characterization of the genome sequences of viruses. In particular, as one can see, the sensitivity of the methods is high and the set of descriptors is adequate for the detection of different serotypes of a virus.

References

[1] D. Bielińska–Waż, T. Clark, P. Waż, W. Nowak, A. Nandy, 2D-dynamic representation of DNA sequences, *Chem. Phys. Lett.* **442** (2007) 140–144.

[2] D. Bielińska–Waż, W. Nowak, P. Waż, A. Nandy, T. Clark, Distribution moments of 2D-graphs as descriptors of DNA sequences, *Chem. Phys. Lett.* **443** (2007) 408–413.

[3] D. Bielińska–Waż, P. Waż, T. Clark, Similarity studies of DNA sequences using genetic methods, *Chem. Phys. Lett.* **445** (2007) 68–73.

[4] P. Waż, D. Bielińska–Waż, 3D-dynamic representation of DNA sequences, *J. Mol. Model.* **20** (2014) #2141.

[5] A. Czerniecka, D. Bielińska–Waż, P. Waż, T. Clark, 20D-dynamic representation of protein sequences, *Genomics* **107** (2016) 16–23.

[6] E. Hamori, J. Ruskin, H Curves, a novel method of representation of nucleotide series especially suited for long DNA sequences, *J. Biol. Chem.* **258** (1983) 1318–1327.

[7] M. A. Gates, Simpler DNA sequence representations, *Nature* **316** (1985) 219–219.

[8] A. Nandy, A new graphical representation and analysis of DNA sequence structure. I: Methodology and application to globin genes, *Curr. Sci.* **66** (1994) 309–314.

[9] P. M. Leong, S. Morgenthaler, Random walk and gap plots of DNA sequences, *Comput. Appl. Biosci.* **11** (1995) 503–507.

[10] M. Randić, M. Vračko, A. Nandy, S. Basak, On 3-D graphical representation of DNA primary sequences and their numerical characterization, *J. Chem. Inf. Comp. Sci.* **40** (2000) 1235–1244.

[11] C. Li, J. Wang, On a 3-D representation of DNA primary sequences, *Comb. Chem. High Throughput Screen.* **7** (2004) 23–27.

[12] Y. Yao, X. Nan, T. Wang, Analysis of similarity/dissimilarity of DNA sequences based on a 3-D graphical representation, *Chem. Phys. Lett.* **411** (2005) 248–255.

[13] Y. Yang, Y. Zhang, M. Jia, C. Li, L. Meng, Non-degenerate graphical representation of DNA sequences and its applications to phylogenetic analysis, *Comb. Chem. High Throughput Screen.* **16** (2013) 585–589.

[14] C. Yuan, B. Liao, T. Wang, New 3D graphical representation of DNA sequences and their numerical characterization, *Chem. Phys. Lett.* **379** (2003) 412–417.

[15] C. T. Zhang, R. Zhang, H. Y. Ou, The Z curve database: a graphic representation of genome sequences, *Bioinformatics* **19** (2003) 593–599.

[16] B. Liao, T. Wang, 3-D graphical representation of DNA sequences and their numerical characterization, *J. Mol. Struct. (Theochem)* **681** (2004) 209–212.

[17] B. Liao, T. Wang, Analysis of similarity/dissimilarity of DNA sequences based on 3-D graphical representation, *Chem. Phys. Lett.* **388** (2004) 195–200.

[18] B. Liao, Y. Zhang, K. Ding, T. J. Wang, Analysis of similarity/dissimilarity of DNA sequences based on a condensed curve representation, *J. Mol. Struct. (Theochem)* **717** (2005) 199–203.

[19] Z. Cao, B. Liao, R. Li, A group of 3D graphical representation of DNA sequences based on dual nucleotides, *Int. J. Quantum Chem.* **108** (2008) 1485–1490.

[20] I. Pesek, J. Žerovnik, A numerical characterization of modified Hamori curve representation of DNA sequences, *MATCH Commun. Math. Comput. Chem.* **60** (2008) 301–312.

[21] W. Chen, B. Liao, X. Xiang, W. Zhu, An improved binary representation of DNA sequences and its applications, *MATCH Commun. Math. Comput. Chem.* **61** (2009) 767–780.

[22] Z. Cao, R. Li, W. Chen, A 3D graphical representation of DNA sequence based on numerical coding method, *Int. J. Quantum Chem.* **110** (2010) 975–985.

[23] J. F. Yu, J. H. Wang, X. Sun, Analysis of similarities/dissimilarities of DNA sequences based on a novel graphical representation, *MATCH Commun. Math. Comput. Chem.* **63** (2010) 493–512.

[24] Y. Li, Y. Qin, X. Zheng, Y. Zhang, Three-unit semicircles curve: A compact 3D graphical representation of DNA sequences based on classifications of nucleotides, *Int. J. Quantum Chem.* **112** (2012) 2330–2335.

[25] N. Jafarzadeh, A. Iranmanesh, C-curve: A novel 3D graphical representation of DNA sequence based on codons, *Math. Biosci.* **241** (2013) 217–224.

[26] Z. Y. Mo, W. Zhu, Y. Sun, Q. L. Xiang, M. Zheng, M. Chen, Z. J. Li, One novel representation of DNA sequence based on the global and local position information, *Sci. Rep.* **8** (2018) #7592.

[27] D. Bielińska–Waż, Graphical and numerical representations of DNA sequences: Statistical aspects of similarity, *J. Math. Chem.* **49** (2011) 2345–2407.

[28] M. Randić, M. Novič, D. Plavšić, Milestones in graphical bioinformatics, *Int. J. Quantum Chem.* **113** (2013) 2413–2446.

[29] M. Mahmoodi–Reihani, F. Abbasitabar, V. Zare–Shahabadi, A novel graphical representation and similarity analysis of protein sequences based on physicochemical properties, *Physica A* **510** (2018) 477–485.

[30] C. Y. Wu, R. Gao, Y. De Marinis, Y. Zhang, A novel model for protein sequence similarity analysis based on spectral radius, *J. Theor. Biol.* **446** (2018) 61–70.

[31] M. Novič, M. Randić, Representation of proteins as walks in 20-D space, *SAR QSAR Environ. Res.* **19** (2008) 317–337.

[32] P. Waż, D. Bielińska–Waż, A. Nandy, Descriptors of 2D-dynamic graphs as a classification tool of DNA sequences, *J. Math. Chem.* **52** (2014) 132–140.

[33] D. Panas, P. Waż, D. Bielińska–Waż, A. Nandy, S. Basak, 2D-dynamic representation of DNA/RNA sequences as a characterization tool of the zika virus genome, *MATCH Commun. Math. Comput. Chem.* **77** (2017) 321–332.

[34] P. Waż, D. Bielińska–Waż, Non-standard similarity/dissimilarity analysis of DNA, *Genomics* **104** (2014) 464–471.

[35] Y. H. Yao, Q. Dai, Ch. Li, P. A. He, X. Y. Nan, Y. Z. Zhang, Analysis of similarity/dissimilarity of protein sequences, *Proteins Struct. Funct. Bioinf. 73* (2008) 864–871.

[36] Y. H. Yao, S. Yan, J. Han, Q. Dai, P. A. He, A novel descriptor of protein sequences and its application, *J. Theor. Biol.* **347** (2014) 109–117.

[37] W. Hou, Q. Pan, M. He, A new graphical representation of protein sequences and its applications, *Physica A* **444** (2016) 996–1002.

[38] V. Aram, A. Iranmanesh, 3D-dynamic representation of DNA sequences, *MATCH Commun. Math. Comput. Chem.* **67** (2012) 809–816.

[39] D. Panas, P. Waż, D. Bielińska–Waż, A. Nandy, S. Basak, An application of the 2D-dynamic representation of DNA/RNA sequences to the prediction of influenza A virus subtypes, *MATCH Commun. Math. Comput. Chem.* **80** (2018) 295–310.

[40] M. G. Guzman, S. B. Halstead, H. Artsob, P. Buchy, J. Farrar, D. J. Gubler, E. Hunsperger, A. Kroeger, H. S. Margolis, E. Martínez, M. B. Nathan, J. L. Pelegrino, C. Simmons, S. Yoksan, R. W. Peeling, Dengue: a continuing global threat, *Nat. Rev. Microbiol.* **8** (2010) 7–16.

[41] T. K. Mackey, B. A. Liang, Threats from emerging and re-emerging neglected tropical diseases (NTDs), *Infect. Ecol. Epidemiol.* **2** (2012) #18667.

[42] S. Bhatt, P. W. Gething, O. J. Brady, J. P. Messina, A. W. Farlow, C. L. Moyes, J. M. Drake, J. S. Brownstein, A. G. Hoen, O. Sankoh, M. F. Myers, D. B. George, T. Jaenisch, G. R. Wint, C. P. Simmons, T. W. Scott, J. J. Farrar, S. I. Hay, The global distribution and burden of dengue, *Nature* **496** (2013) 504–507.

[43] R. de Alwis, K. L. Williams, M. A. Schmid, C. Y. Lai, B. Patel, S. A. Smith, J. E. Crowe, W. K. Wang, E. Harris, A. M. de Silva, Dengue viruses are enhanced by distinct populations of serotype cross–reactive antibodies in human immune sera, *PLoS Pathog* **10** (2014) #e1004386.